



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Zhang, Ligang, [Tjondronegoro, Dian W.](#), & [Chandran, Vinod](#)
(2014)

Random Gabor based templates for facial expression recognition in images with facial occlusion.

Neurocomputing, 145, pp. 451-464.

This file was downloaded from: <https://eprints.qut.edu.au/71005/>

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<https://doi.org/10.1016/j.neucom.2014.05.008>

Random Gabor based templates for facial expression recognition in images with facial occlusion

Ligang Zhang, Dian Tjondronegoro and Vinod Chandran

Science and Engineering Faculty, Queensland University of Technology

2 George Street, Brisbane, QLD 4000, Australia

{L44.zhang, dian, v.chandran}@qut.edu.au

Abstract Robust facial expression recognition (FER) under occluded face conditions is challenging. It requires robust algorithms of feature extraction and investigations into the effects of different types of occlusion on the recognition performance to gain insight. Previous FER studies in this area have been limited. They have spanned recovery strategies for loss of local texture information and testing limited to only a few types of occlusion and predominantly a matched train-test strategy. This paper proposes a robust approach that employs a Monte Carlo algorithm to extract a set of Gabor based part-face templates from gallery images and converts these templates into template match distance features. The resulting feature vectors are robust to occlusion because occluded parts are covered by some but not all of the random templates. The method is evaluated using facial images with occluded regions around the eyes and the mouth, randomly placed occlusion patches of different sizes, and near-realistic occlusion of eyes with clear and solid glasses. Both matched and mis-matched train and test strategies are adopted to analyze the effects of such occlusion. Overall recognition performance and the performance for each facial expression are investigated. Experimental results on the Cohn-Kanade and JAFFE databases demonstrate the high robustness and fast processing speed of our approach, and provide useful insight into the effects of occlusion on FER. The results on the parameter sensitivity demonstrate a certain level of robustness of the approach to changes in the orientation and scale of Gabor filters, the size of templates, and occlusions ratios. Performance comparisons with previous approaches show that the proposed method is more robust to occlusion with lower reductions in accuracy from occlusion of eyes or mouth.

Keywords *Facial expression recognition, Face occlusion, Gabor based template, Support vector machine*

1 Introduction

Facial expression recognition (FER) has received significant interest from computer scientists and psychologists over recent decades, largely driven by its potential in applications such as human computer interaction, video surveillance, and multimedia content analysis. Although many systems for FER [1] have been proposed and implemented, the vast majority of them use non-

occluded facial images taken under controlled laboratory conditions [1]. In the real world, however, facial occlusion is common and arises mainly from clothing, eyewear, and hand movements in front of a face. Building robust FER systems under facial occlusion is still a challenging task.

There are two types of facial occlusion: temporary and systematic [2]. Temporary occlusion can result from a part of the face obscured by a hand or an object or owing to head movement. Systematic occlusion results from wearing an item such as sunglasses, a scarf or mask. Unlike other degradations such as pose variations, whose characteristics can be inferred beforehand, facial occlusion is particularly difficult to handle due to its “random” characteristic: occlusion can occur at random positions and occluding patches can be arbitrarily large in size.

Existing FER studies on handling occlusion try to reconstruct the occluded geometric or texture features based on the configuration and visual properties of human face. Techniques for reconstructing geometry include principal component analysis (PCA) [2], the improved Kanade-Lucas tracker [3], [4], Bayesian tracker [5], and the transferable belief model [6], and those for reconstructing texture rely on robust PCA [7], [8]. These systems depend heavily on accurate detection of geometric points and occlusion regions. They either lack the capacity to fully capture the regional texture features or lose a part of local texture information due to the reconstruction error. Local texture information is important for FER.

Investigations into the effect of occlusion on recognition performance have either used a combination of geometric and texture features [9], [10], or approached FER from the view of human perception [11], [12]. While these works have provided useful insight such as the most discriminative facial areas for specific expressions, they have been restricted to occlusion over specified regions such as the mouth, eyes, nose, or left and right sides. To model the “random” characteristic of occlusion in the real world, it is desirable to test with occlusion at random locations and occluding patches of different sizes. Further, most training and testing data sets used in these studies were matched and have been both occluded in the same manner. It is of interest to test the recognition performance on occluded images when non-occluded images are used for training (mismatched conditions). This is particularly important for real-world applications in that it can eliminate the step of generating the occluded training images by requesting clients at the time of enrolment. It is not practically possible to obtain training images with all possible types of occlusion.

In this work, a new approach is adopted and there is no attempt to reconstruct occluded parts of a face. Instead, a robust FER approach using randomly sampled Gabor-based templates is proposed. The templates are extracted by a Monte Carlo algorithm and serve as a pool of local features; therefore, only a part of them are influenced by occlusion. Template matching is used to find *the most similar features* located within a space *around* these templates. In this way, robust features are generated in the sense that occluded templates can be replaced by nearby non-occluded templates during matching. The salient features that are least influenced by occlusion are further determined in the learning process using a norm-based support vector machine (SVM). This approach is designed based on the observation that even when parts of a face are occluded other parts still contain facial expression information, and a machine learning algorithm can learn to avoid the occlusion from training samples automatically in a feature selection process. This

approach is inspired by recent work on object recognition [13] and action recognition [14], for which Gabor based templates exhibit excellent performance in terms of position, scale and orientation-invariance, and randomly extracted Gabor features show a good generalization capacity of representing all features. It should note that Gabor filters are one of the most widely used feature descriptors in previous FER on occluded data [9], [10]. As most cases in realistic situations, it is assumed that the size of occlusion is limited to a certain proportion of the face region (i.e. less than 40%) and the location of occlusion can be fixed or random in the train-test datasets. The adoption of a space for template matching also determines that the proposed approach is more suitable for small occlusion rather than large ones.

Compared with previous studies, the novelty of this paper primarily consists of two parts: a) a novel FER approach that is based on the Monte Carlo based feature selection and template matching, which are specifically designed to overcoming facial occlusion by utilizing un-occluded information in neighbouring regions; and b) one of the most comprehensive and extensive evaluations of FER under occlusion conditions to date, which include five types of patch-like occlusion (occluded eyes, occluded mouth, and three occluding patches at random locations and with different sizes), two types of near-realistic occlusion (clear and solid glasses), and two testing strategies (training and testing are both performed on occluded images i.e. matched; training is performed on non-occluded images and testing is done on occluded images i.e. mismatched). To the best of our knowledge, this is the first study that investigates the effects of random occlusion patches on the overall performance and performance of each expression, and uses the two testing strategies. In our previous work [15], the classification results of the proposed approach on the CK database has been reported, and this paper presents an extension and continuation of [15] with a) detailed description about the proposed approach, b) thorough review on previous related work, c) expanded experiments of different occlusion types on the JAFFE database, d) sensitivity analysis of vital parameters, e) inclusion of occlusion of eyes with clear and solid glasses, and f) computational time performance.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed approach. Section 4 describes the experiments. Finally, conclusions are drawn in Section 5.

2 Related work

Existing FER studies on occluded data can be generally classified into two categories: a) methods that address occlusion by reconstruction, and b) investigations into the effects of occlusion on recognition performance.

2.1 Methods for handling occlusion

Occlusion handling methods try to remove the impact of occlusion by recovering missing geometric or texture features based on the configuration and visual properties of the face. Bourel *et al.* [3] conducted the first study on handling occlusion of the mouth, upper face, and left/right half of the face for FER using recovered geometric features. An enhanced Kanade-Lucas tracker is

employed to recover lost or drifting facial points. Independent local spatio-temporal vectors are created from geometrical relations between 12 facial points. A rank-weighted KNN local classifier is then used to make sure that occlusion in a region does not affect other regions. To further reduce noise in the spatio-temporal vectors, they have added a state-based model to convert continuous values into discrete ones through a quantization process [4]. Towner and Slater [2] utilized three PCA-based methods to reconstruct the positions of missing points at the top and bottom of the face. Their results showed that occlusion of the top of the face can be reconstructed with little loss, while occlusion of the bottom is reconstructed less accurately but still gives comparable expression recognition performance. Dat and Ranganath [5] proposed a Bayesian tracker. This tracker augments the KLT tracker with a Bayesian feedback mechanism to improve results of tracking facial points under temporal occlusion by hands. Hammal *et al.* [6] recognized facial expressions from partially occluded images based on facial point deformations and a modified transferable belief model (TBM). The different types of occlusion are simulated by adding *bubble* masks into the face and handled by the TBM, which can integrate information from different local facial regions, and deal with uncertain and imprecise data. Miyakoshi and Kato [16] presented an emotion system that utilizes a Bayesian network classifier to infer the dependencies among target attribute and explanatory variables without filling in the facial gap caused by occlusion.

Techniques have also been used for recovering texture features in occluded regions. Xia *et al.* [7] first detected regions of occlusion based on robust PCA (RPCA) and saliency detection, and then replaced occluded regions by corresponding regions in the face reconstructed using RPCA. A similar RPCA algorithm was also applied for reconstructing the regions of occluded eyes and mouth in [8]. Different from all above work, Cotter [17], [18] represented occluded test images using a linear combination of pixel vectors from local regions in training samples from the same class. Yongmian and Qiang [19] inferred facial expressions in occluded frames via temporal reasoning. Recently, Huang *et al.* [20] proposed a complete system consisting of facial representation, occlusion detection, and multiple feature fusion in video sequences. Temporal texture and shape features from the eyes, nose, and mouth components are integrated using a mean rule and a weight learning strategy, and the occluded components in sequences are determined using a binary codebook generated based on both sparse representation and residual statistics of these components.

FER performance improvements have been provided by these methods, but they still have the following shortcomings:

- 1) Limited robustness against occlusion. Techniques for constructing geometric features face the difficulty to accurately detect and track the geometric points in video with occluded regions. Techniques for constructing texture features have to pre-locate occluded regions, which is still a challenging issue.

- 2) Loss of local texture information. The PCA-based recovering technologies cannot fully represent complicated and local changes of facial appearance without some loss of useful texture information.

To overcome these shortcomings, the random template method proposed in this paper does not require pre-location of occluded regions and does not depend on recovering texture in the

occluded regions. It uses randomly collected features and selects a subset of those that is least influenced by the types of occlusion present in the training set.

2.2 Investigations into the effects on performance

Investigations into the effects of occlusion on FER performance are generally based on texture features in static images. Buciu *et al.* [9] explored the classification performance of seven emotions under occlusion of the mouth and occlusion of the eyes using Gabor features. Their results showed that the eyes and mouth affect the overall performance on the JAFFE database in a similar manner, while the mouth plays a more important role than the eyes on the CK database. The eyes and mouth have the most impact on sad and neutral expressions, respectively, on JAFFE, while they both have the most impact on anger and sadness on CK. Buciu *et al.* [10] also presented an analysis, probably the most detailed to date, on the effects of occlusion of the eyes, mouth, and left and right face regions, on the FER performance. Their results demonstrate that occlusion on the left and right do not affect the recognition accuracy, but occlusion of the mouth affects anger, fear, happiness and sadness, while occlusion of the eye affects disgust and surprise, more than other emotions.

Attempts have also been made to experimentally analyze the effect of occlusion on facial expressions based on human perception. Halliday [21] demonstrated that the mouth and eyes are the most important regions for accurately identifying genuine happiness, sadness and fear. Bassili [11] indicated that the bottom part of the face tends to produce a better overall recognition performance than the top part. Nusseck [12] reported that the combination of eyes, eyebrows, and mouth regions is sufficient to produce acceptable recognition rates.

Useful insights have been provided by these studies, but they have:

- 1) Used a limited range of occlusion types. Only occluded regions over the mouth, eyes, nose, and left/right halves of the face are used for testing. In the real-world, a face can become occluded at any location, and it is necessary to test more comprehensively with occluded regions at random locations and of random sizes.

- 2) Adopted a matched train-test strategy. Both the training and test sets are occluded images. It is also important to test the mismatched strategy, where no-occluded images are used for training and occluded images are used for testing.

Overcoming these limitations, the evaluation in this paper addresses the effects of a wider range of occluded regions on FER performance, and uses a mismatched train-test strategy as well.

3 The proposed approach

Figure 1 demonstrates the framework of the proposed approach. For an input image, the facial region is located by the widely used Viola-Jones face detector and scaled to 48×48 pixels. Occluded regions are simulated by adding masks into the facial regions. Then Gabor images are obtained by convolving eight-scale and four-orientation Gabor filters with the facial regions. A Monte Carlo algorithm is used to extract a set of templates, which are only partially influenced by occlusion. Template matching is then performed to find the most similar features located within a

space around these templates, resulting in distance features robust against occlusion. Based on the distance features, a norm-based support vector machine (SVM) is used for selecting salient templates that are least influenced by occlusion. At the testing stage, the same template matching is performed on the salient templates to obtain the distance features in a testing image, which are fed into an SVM to recognize seven emotions: anger (AN), disgust (DI), fear (FE), happiness (HA), neutral (NE), sadness (SA) and surprise (SU).

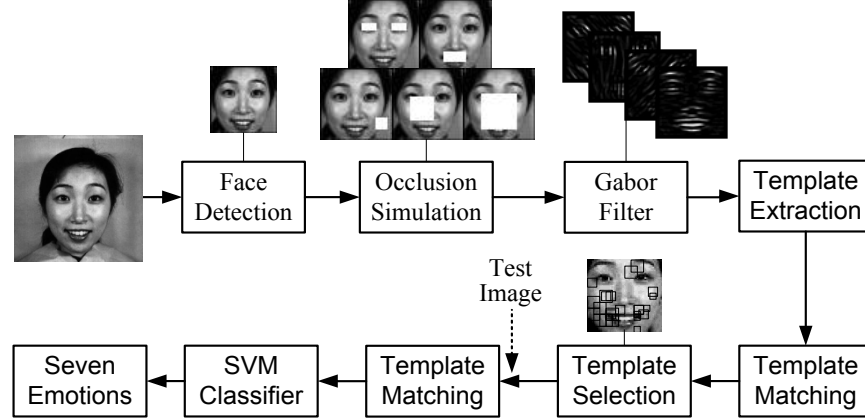


Fig. 1 Framework of the proposed approach

3.1 Gabor filters

Gabor filters are one of the most commonly used texture descriptors for representing facial expression information under occlusion situations in previous studies (see Section 2). In this paper, 2D Gabor filter [22] is adopted and it can be mathematically expressed as:

$$F(x, y) = \exp\left(-\frac{(X^2 + \gamma^2 Y^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right)$$

$$X = x \cos \theta + y \sin \theta; Y = -x \sin \theta + y \cos \theta \quad (1)$$

where, orientation θ , the effective width σ , the wavelength λ , the aspect ratio γ . Instead of the widely used five scales, eight scales (5:2:19 pixels) are adopted here to test the results using a larger number of scales. The values of the rest of the parameters are set based on [13], which reported high performance on object recognition using these parameters. As a result, four orientations (-45° , 90° , 45° , 0°) are used. The given image is effectively convolved with impulse response of the Gabor filters as in equation 1, resulting in a series of Gabor images with features such as bars and edges usefully emphasized for better discriminating between facial expressions.

3.2 Template extraction

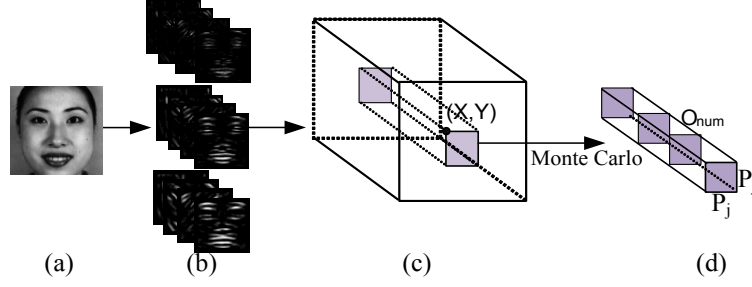


Fig. 2 (a) Input image, (b) set of Gabor filter outputs, (c) one scale from the space of all Gabor outputs showing a template being selected at random location (X, Y) , (d) the selected 3D template of width P_j , height P_j and O_{num} orientations.

As shown in Figure 2, template extraction adopts a Monte Carlo algorithm to collect a set of templates from the occluded Gabor images. Each template has patches of Gabor filter outputs at four orientations, all from the same scale. These templates serve as a pool of local features and contain emotional information robust against occlusion. Figure 3 explains the algorithm. For each emotion and each template size, a number (Num) of templates are extracted. The final template set is obtained by combining the Num extracted templates of all emotions and all template sizes. To extract each of the Num templates, a five-step process is proposed: First, randomly select one image I from all training images with the emotion E_k . Second, randomly select one Gabor scale S_m from eight scales of the image I . Third, randomly select the X- and Y-axis positions from the scale S_m . Fourth, extract one template with a size of $P_j \times P_j \times O_{num}$ at the position of X and Y in the scale S_m . Finally, record the matching area and matching scale (details explained in Section 3.3).

This algorithm is designed based on the “random” characteristic of regions of facial occlusion and the fact that local features have an advantage over holistic features in overcoming occlusion [4]. In real situations, facial occlusion can randomly occur across the whole face region, but often influence only on a small part of the region. Therefore, it is expected that only a small proportion of the randomly extracted templates are affected by facial occlusion, while the majority of the templates obtained in the non-occluded regions remain uninfluenced. Similar algorithms have been used to extract a subset of features with a good generalization capacity of representing all features [13]. In this paper, four template sizes are used: $2 \times 2 \times 4$, $4 \times 4 \times 4$, $6 \times 6 \times 4$, $8 \times 8 \times 4$, and the Num is set to 1000. Given seven emotions, the final set contains 28,000 templates.

Input: Template size P_j ($j=1, \dots, 4$); emotion E_k ($k=1, \dots, 7$); Gabor scale S_m ($m=1, \dots, 8$); Gabor orientation number O_{num} .

Output: Templates, matching area and matching scale.

For each emotion E_k and each template size P_j

For 1 to Num ($Num = 1000$)

 Randomly select one image I from training images of E_k ;

 Randomly select one scale S_m from all scales in I ;

 Randomly select X- and Y-axis positions in S_m ;

 Extract templates with sizes of $P_j \times P_j \times O_{num}$;

 Record the matching area $Area$;

Record the matching scale S_m ;

End

End

Return *Templates, matching area and matching scale.*

Fig. 3 Algorithm for template extraction

3.3 Template matching

Template matching aims to find the most similar features located within a space around the extracted templates. The matching helps to reduce the influence of occlusion because occluded templates can be replaced by nearby non-occluded templates. Thereby, robust features can be extracted. For each of the extracted templates, the matching produces a distance value by performing the following two steps, and the distances of all templates are concatenated to form the final feature set.

First, the matching area and matching scale are defined for each template to provide the matching space. As shown in Figure 4, the matching area has twice the width and height of each template but with the same orientation number and centre point. The matching scale is set as the same with the scale of P_a because the facial regions obtained by the Viola-Jones detector normally have the same scale. A larger matching area (e.g. the whole face) or matching scale can be used, but it also may increase the possibility of error introduced because the extracted template is matched against more templates which may contain similar texture.

Second, the distances between the template P_a and all templates within the matching space are calculated. Each calculation takes two templates as inputs and yields one value based on a distance metric. The minimum value in the matching space is the chosen feature for P_a . Two distance metrics are used: sparse L_1 (SL_1) and sparse L_2 (SL_2). The sparse metrics calculate the distances using the maximum value of all orientations in the template:

$$SL_1 : \|P_b - P_c\| = \frac{1}{P_j * P_j} \sum_{i=1}^{P_j} \sum_{j=1}^{P_j} (\max_{o=1}^{O_{num}} \sum_{o=1}^{O_{num}} P_b^{ijo} - \max_{o=1}^{O_{num}} \sum_{o=1}^{O_{num}} P_c^{ijo}) \quad (2)$$

$$SL_2 : \|P_b - P_c\| = \frac{1}{P_j * P_j} \sqrt{\sum_{i=1}^{P_j} \sum_{j=1}^{P_j} (\max_{o=1}^{O_{num}} \sum_{o=1}^{O_{num}} P_b^{ijo} - \max_{o=1}^{O_{num}} \sum_{o=1}^{O_{num}} P_c^{ijo})^2} \quad (3)$$

where, P_b and P_c represent two templates, P^{ijo} represents the pixel value in the i^{th} row, j^{th} column and o^{th} orientation of the templates, O_{num} indicates the number of Gabor orientations, and P_j indicates the size of the j^{th} template.

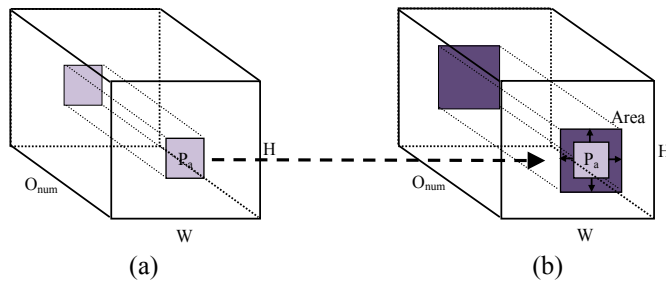


Fig. 4 Matching area. (a) One template P_a and its (b) corresponding matching area ‘Area’ in a new image. W and H are the width and height of images.

3.4 Template selection

Norm-based SVM [23] is used to select a subset of the most discriminative templates and eliminate those influenced most by occlusion. The SVM solves classification tasks by finding a hyperplane in a high dimensional space that can separate the negative examples from the positive examples with a maximum margin. The output prediction P for a linear SVM is expressed as:

$$P(x) = \text{sgn}(w^T x + b) = \text{sgn}(\sum_j w_j x_j + b) \quad (4)$$

where, x_j is the j^{th} distance obtained from template matching, which is concatenated to form a feature vector, and w_j is its weight. Thus, a feature x_j with the weight w_j close to 0 has smaller impact on the prediction result than features with big absolute values of weights. The weights w can be seen as normal vectors to the hyperplane determined by the SVM classifier, and these normal vectors are updated every round to reflect the importance of each feature. Template selection with the SVM will also reduce computation time.

In this paper, a one-against-all SVM is trained for each emotion. For each template, the average weights over the seven emotions are updated each training round based on the trained SVM. The templates with low average weights are dropped out (i.e. setting the weights to 0). The number of the selected templates is controlled in such a way that it initially starts from 28,000, and reduces to a half at each training iteration until it becomes less than 1000, then decreases by 50 until less than 100, and then by 10 when bigger than 10.

4 Experimental results

This section introduces the databases and simulation of occluded regions, and analyzes the effects of different types of occlusion on the overall FER performance and the recognition performance for each expression. Classification accuracies and average computational time are reported and performance is compared with previous approaches.

The effects of occlusion on the performance are described based on two train-test strategies: a) *matched strategy* means that training and testing are both performed on occluded images; b) *mismatched strategy* indicates that training is performed on non-occluded images and testing is done on occluded images. The average correct recognition accuracy over 10 cross validations with a linear SVM classifier is used. The performance for the case of no occlusion (No) serves as a baseline for comparisons. Since SL_1 and SL_2 produce similar performance for the two testing strategies and for most types of occlusion, the performance obtained by SL_1 alone is used for analysis of the results.

4.1 Databases

The Cohn-Kanade (CK) database [24] is one of the most comprehensive databases for facial expression benchmark tests. The released part includes 486 digitized image sequences from 97 subjects. These sequences from neutral to target display were digitized into 640×480 or 640×490 pixels. Six basic expressions were based on descriptions of prototypic emotions. For the experiment here, 1,616 images that represent one of the seven emotions are selected from 92 subjects. For six basic emotions, the four frames with the highest intensity of emotions are chosen in each sequence. For the neutral case, the first frame from each video sequence is chosen. At most four are retained, ensuring that the neutral class has almost the same number of images as the other emotion classes for each subject. Only when there are less than four emotions for a subject is there a discrepancy. The statistics of the chosen CK images are listed in Table 1. All images are classified into 10 similar sets and all images of one subject are included in the same set.

Table 1 Statistics of images chosen from CK database

	Total	AN	DI	FE	HA	NE	SA	SU
Image	1616	132	140	200	316	356	184	288
Subject	92	33	35	50	79	92	46	72

The Japanese female facial expression (JAFPE) database [25] contains 213 gray images of seven facial expressions posed by 10 Japanese females. Each image has a resolution of 256×256 pixels. Each object has three or four frontal face images for each expression. All images have been rated on six emotion adjectives by 60 subjects. All images are used here and classified into 10 sets according to the subject identity.

4.2 Simulation of occlusion

No comprehensive public FER database has been reported with images that exhibit different types and extent of facial occlusion. To simulate such occlusion, we add white masks to different positions of the face region, including (i) the two eyes, (ii) the mouth, (iii) randomly placed 8×8 patch R_8 , (iv) R_{16} and (v) R_{24} . R_S means adding a mask with a size of $S \times S$ into the face region at a random location. This random patch technique has been used previously to simulate occluded regions in face recognition [26]. A sample set of occluded facial images on the CK database is shown in Figure 5 (a). In addition to the five types of patch-like occlusion, two types of near-realistic occlusion are also included: by clear and solid glasses, which are simulated by adding a white frame and by adding a black frame together with a white interior over the eye region, respectively, as shown in Figure 5 (b). The motivation for testing with these types of occlusion is the fact that in real situations glasses often occlude the two eyes, scarves and masks often occlude the mouth, and other unpredictable objects can occlude any part of the face randomly. In addition, eyes and mouth are widely regarded as the most important areas for facial expressions, therefore, occlusion of the eyes and mouth can be used to validate the robustness of the proposed approach. It is worth mentioning that occlusion of the left or right halves of the face is not tested here as this type of occlusion has been shown to have little effect on the performance [10].

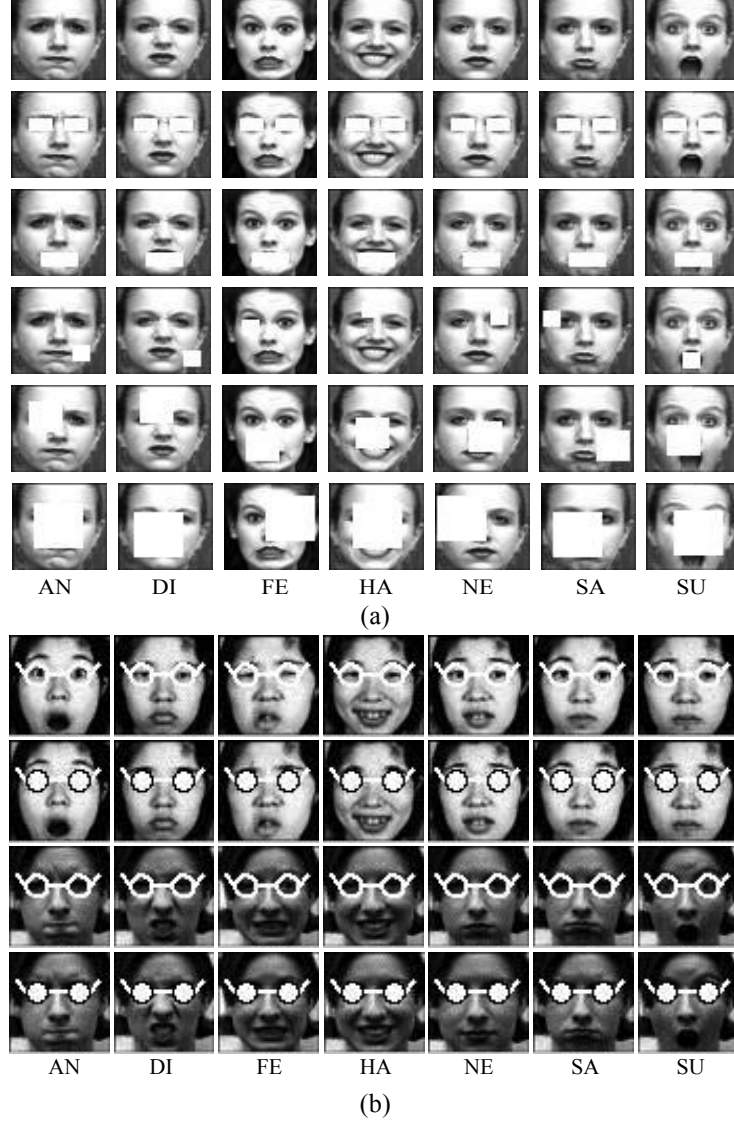


Fig. 5 Samples of different types of simulated occlusion. (a) From top to bottom: no occlusion, occluded eyes, occluded mouth, R_8 , R_{16} and R_{24} on the CK database. (b) Near-realistic occlusion by simulated clear and solid glasses on the JAFFE (first two rows) and CK (last two rows) databases.

4.3 Effect of occlusion on overall FER performance

Figure 6 demonstrates the effects of the five types of patch-like occlusion on the overall FER performance of the proposed approach.

(1) Matched train-test strategy (using occluded images for training)

For both the CK and JAFFE databases, no occlusion (No) achieves the highest overall performance. The second best performance is for occluded eyes with 95.1% and 80.3% accuracy on the CK and JAFFE databases respectively. The next ones are R_8 , occluded mouth and R_{16} , whereas R_{24} ranks last. Compared to no occlusion, occlusion of the eyes leads to a small performance reduction (<1% for both databases), whereas occlusion of the mouth results in reductions of about 5% for CK and 3% for JAFFE. When the size of random occlusion patches increases from 8×8 to 24×24 , the overall performance has a significant reduction. However, the proposed approach still has 75% accuracy on CK and 48.8% accuracy on JAFFE under R_{24} , for

which one-quarter of the face is occluded. This demonstrates the high robustness of the proposed approach.

The proposed approach has much lower accuracy on JAFFE than on CK for all types of occlusion, which is owing to the more ‘exaggerated’ emotions in CK than in JAFFE. R_{16} and R_{24} have bigger impact on the performance on JAFFE than that on CK. This is partially caused by the fact JAFFE images need bigger sizes of Gabor templates than CK images to represent the useful information for FER as shown in [27]. Accordingly, larger sizes of random occlusion cause higher variance in the Gabor templates extracted from JAFFE images during training.

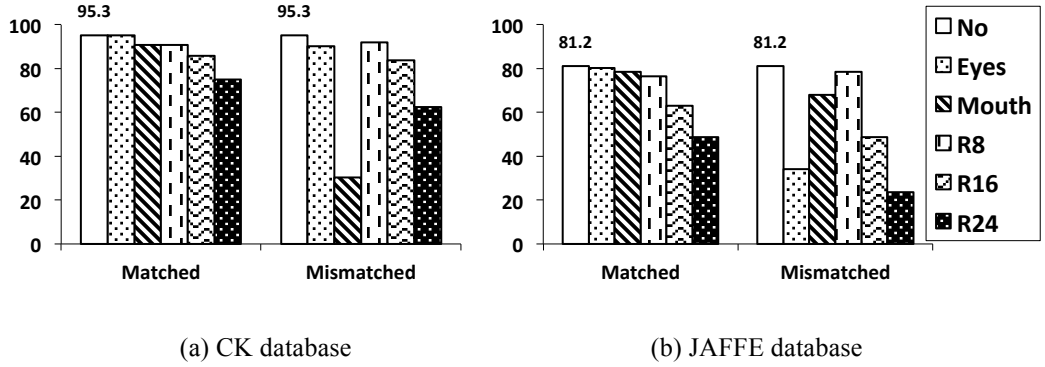


Fig. 6 Overall recognition performance (%) under five types of occlusion on the (a) CK and (b) JAFFE databases. “No” indicates no occlusion. Under the matched strategy, R_8 has the biggest impact on overall performance for both the databases. Under the mismatched strategy, occlusion of the mouth degrades performance on CK significantly, while occlusion of the eyes and R_{24} exert major impact on performance on JAFFE.

(b) Mismatched train-test strategy (using non-occluded images for training)

For both the databases, the best and second performers are no occlusion and R_8 respectively. An increase in the size of random occlusion patches results in lower performance. On the other hand, there are also some differences in performance on the two databases. Occlusion of the mouth has a big impact on the performance on CK, but little impact on JAFFE. Occlusion of the eyes and R_{24} significantly influence the performance on JAFFE, but not so much for CK. This shows that the most influencing types of occlusion may be different for different data. Emotions in JAFFE are expressed mainly through regions of the two eyes, while those in CK are largely through the mouth region. The results agree with previous finding [27] that the Gabor features selected by Adaboost from JAFFE images focus more on the eyes, while those from CK images are more on the mouth. Such differences can arise from factors such as cultural differences, the intensity of posed emotions, etc.

The performance for mismatched train-test strategy is significantly reduced from that for a matched train-test strategy for most types of occlusion. For the CK database, occlusion of the mouth leads to the biggest reduction (e.g. from 90.8% to 30.3%). The one that follows in performance reduction is R_{24} , which suffers from a reduction from 75.0% to 62.5%. Occlusion of the eyes and R_{16} lead to reductions of about 5% and 2% only, respectively. Surprisingly, R_8 achieves little higher accuracy than that using matched train-test strategy, which may be regarded

as being within a reasonable performance fluctuation. For the JAFFE database, occlusion of the eyes brings about the biggest reduction (e.g. from 80.3% to 34.3%), which is followed by R_{24} (e.g. from 48.8% to 23.5%). In contrast, R_8 even has little higher accuracy, which is similar to the result on CK. This indicates that the use of a matched or mismatched train-test strategy has little impact on R_8 .

As shown by the above results, for both the testing strategies and both the databases, R_8 has little influence on the overall performance, while R_{24} exerts a great influence. This implies that the *size of the occluded regions* is one important factor determining the performance. Our results also reveal that *regions of the two eyes* have a big effect on JAFFE, while the *mouth region* has a significant impact on CK. The performance difference between the two testing strategies is owing to the fact that the SVM feature selector only selects features from non-occluded regions based on the knowledge learnt from training images during matched conditions, whereas there is no such knowledge when non-occluded images are used for training in the mismatched case.

4.4 Effects of occlusion on each facial expression

(1) Matched train-test strategy

Figure 7 illustrates the effects of the five types of occlusion on each expression based on the matched train-test strategy. For both the databases, occlusion of the eyes has little influence on any expression, except for anger and neutral on JAFFE. Furthermore, it slightly outperforms no occlusion (No) for anger, fear and sadness on CK and fear on JAFFE. This is due to the performance fluctuations caused by the Monte Carlo algorithm of template extraction. Occlusion of the mouth tends to impact different emotions on the two databases. For CK, occluded mouth has a relatively significant effect on anger, sadness and fear, while it has little effect on happiness, neutral and surprise. This accords with the results in [9], and it is reasonable for anger, sadness, and fear, but is contrary to the common understanding that the most distinguishing characteristic of surprise is an “open mouth”. The reason may be that surprise on faces in the CK database is expressed not only by the mouth, but also by other regions. For JAFFE, occluded mouth impacts happiness the most, following by neutral and sadness.

Increasing size of random occluded patches decreases the performance for most of the expressions, except for happiness on JAFFE, which has higher accuracy under R_{16} than under R_8 . All expressions are most affected by R_{24} , and this affection is more obvious on JAFFE than CK. Furthermore, anger and sadness on CK, and disgust, neutral, surprise and sadness on JAFFE are also strongly affected by R_{16} . For both the databases, similar performance is obtained for disgust and surprise with occlusion of the eyes, the mouth and occlusion by random R_8 patches. This is also true for happiness and neutral on CK. In addition, neutral and surprise also have the same accuracy under R_{16} and R_{24} on JAFFE. It should be noted that, due to the Monte Carlo algorithm of template extraction and varying locations of simulated occlusion regions, a larger random occlusion can sometimes lead to better performance than smaller occlusion for certain emotions. For example, R_{16} obtains higher accuracy than R_8 for happiness on the JAFFE database. This is also true for some results of the mismatched train-test strategy in Figure 8.

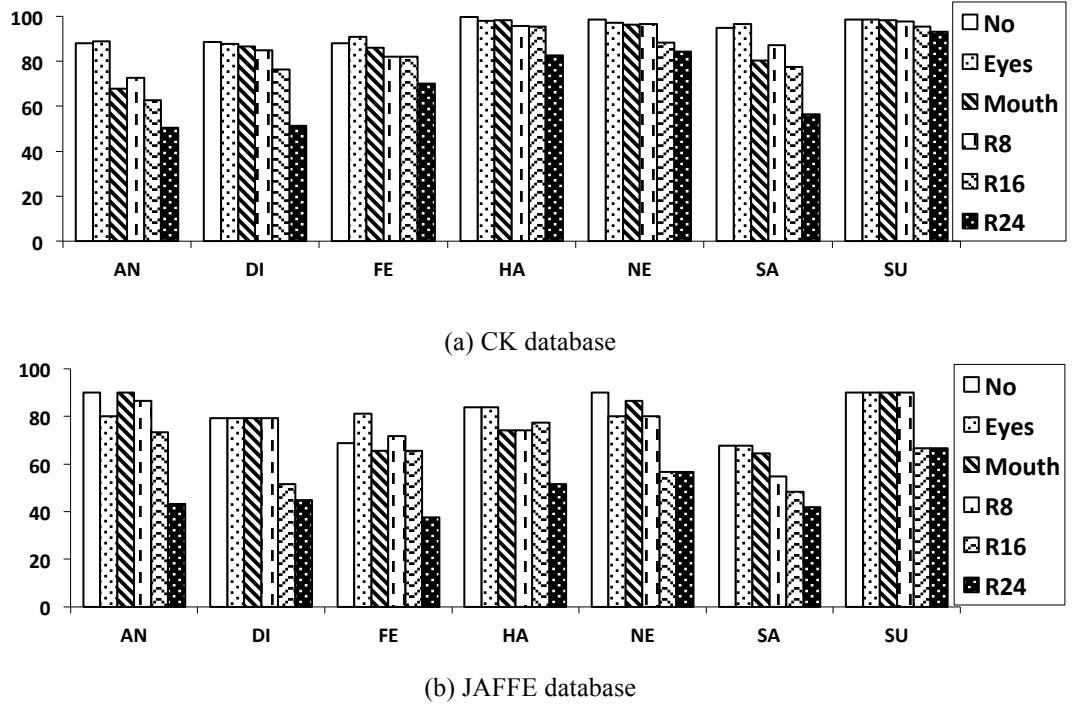


Fig. 7 Effects of five types of occlusion on the recognition performance (%) of each facial expression using the matched train-test strategy. Note that R_{24} impacts all emotions significantly on the two databases.

(2) Mismatched train-test strategy

Figure 8 (a) shows the effects on each expression for the CK database based on the mismatched strategy. Occlusion of the mouth demonstrates a significant impact on the performance of sadness, neutral, disgust, surprise, and fear, but a small effect on that of anger and happiness. All sad images are misclassified, and this agrees with the result in [3] that only 20% accuracy is obtained for sadness. Moreover, neutral, disgust and surprise have far lower accuracy at 3.8%, 15%, and 15% respectively, than anger and happiness. Occlusion of the eyes has little influence on all emotions. These findings confirm the results in [9] that the eyes region is not as important as the mouth region for FER on the CK database. R_{24} and R_{16} also have a big influence on all emotions, while R_8 has little impact. Note that for anger higher accuracy is obtained under R_{24} than under R_{16} (62.6% versus 54.6%) and this is an exception.

Figure 8 (b) illustrates the effects on the JAFFE database. R_8 has little impact on any emotion, while R_{24} has a big impact on all emotions. R_{16} also tends to have a significant effect on most of the emotions. However, the effects of R_8 and R_{16} on JAFFE are not as significant as those on CK. For both the databases, R_{16} has little influence on performance for fear. On the CK database occlusion of the mouth has a significant impact on performance. By contrast, on JAFFE, occlusion of the mouth exerts a relatively small influence on most emotions. Occlusion of the eyes exerts the biggest influence amongst all types of occlusion, on anger, neutral, and surprise. This illustrates that the eyes rather than the mouth convey the most discriminative information in the JAFFE images, as opposed to the results in the CK images. Note that for fear and disgust accuracy is better under R_8 and R_{16} than for the case of no occlusion.

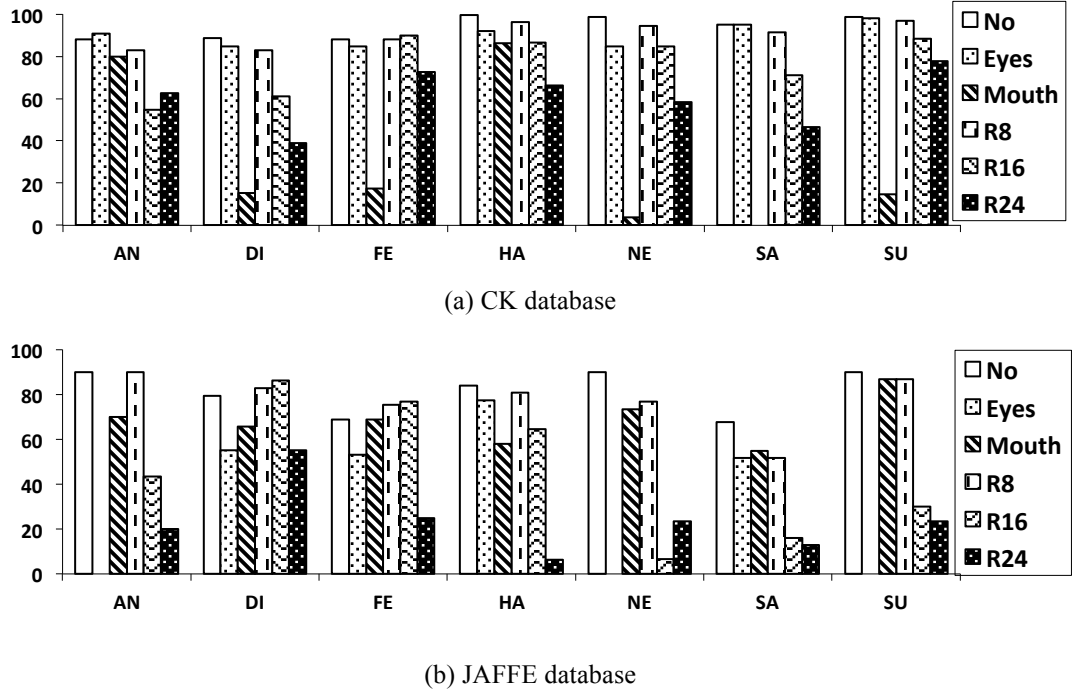


Fig. 8 Effects of five types of occlusion on the recognition performance (%) of each facial expression using the mismatched train-test strategy. Note that occlusion of the mouth degrades recognition performance of disgust, fear, neutral, sadness and surprise significantly but does not affect anger and happiness as much on the CK database. For the JAFFE database, occlusion of the eyes significantly impacts performance of anger, neutral, and surprise, but does not affect happiness much, while R_{24} has the biggest impact on fear, happiness and sadness.

Compared to the results using a matched strategy in Figure 7, R_{24} and R_{16} result in large performance reductions for both the databases. In addition, occlusion of the mouth and occlusion of the eyes also degrade the performance significantly for CK and JAFFE, respectively. On the other hand, R_8 leads to similar results for the two testing strategies. When matched train-test conditions are used, all emotions are mostly affected by R_{24} for both the databases. For mismatched conditions, all emotions except for anger and happiness are mostly affected by occluded mouth on CK, and anger, neutral, and surprise are mostly affected by occluded eyes, while fear, happiness and sadness are mostly affected by R_{24} on JAFFE. Similar to the results presented in [9], [10], occlusion of the mouth is observed to result in a bigger performance reduction than occlusion of the eyes for most expressions on the CK database.

4.5 Parameter sensitivity

The relevant parameters in the proposed approach include: (a) the orientation and scale of Gabor filters, (b) the size of templates, (c) the ratio of the height and width of occlusion, (d) the size of the match area, and (e) the distance function. Changes made to these parameters may have big impact on the performance, thus, it would be interesting to investigate the sensitivity of the approach to parameters that use values different from those set initially. Due to space limitation, this section reports only the results obtained using different values for each of the parameters of (a-

c) and the database used is JAFFE.

(1) Gabor filter

The orientation and scale of Gabor filters imposed on facial images are two key parameters that determine the effectiveness of the extracted texture features. Fig. 9 compares the classification results obtained using two, four, and eight orientations of Gabor filters (the number of the scale of Gabor filters is fixed to eight). The three sets of orientation are $(90^\circ, 0^\circ)$, $(-45^\circ, 90^\circ, 45^\circ, 0^\circ)$ and $(90^\circ, 67.5^\circ, 45^\circ, 22.5^\circ, 0^\circ, -22.5^\circ, -45^\circ, -67.5^\circ)$ respectively. It can be seen that using the default four orientations of Gabor filters produces the highest accuracy among the three sets of orientation values on all types of occlusion under matched train-test conditions, and it is also the best performer for non-occlusion and occlusions of R_8 , R_{16} , and R_{32} under mismatched conditions. In addition, an increase of the orientation of Gabor filters from four to eight leads to slightly inferior performance and both of them outperform using two orientations (more than 5%) for nearly all occlusions and both mismatched and matched conditions. The result may indicate that Gabor filters require at least four orientations to be capable to capture most of the information for discriminating facial expressions in occluded images. A significantly higher accuracy is reported for using eight orientations over using two and four orientations for occlusion of the mouth under mismatched conditions, implying that it might be helpful to adopt a larger number of the orientation of Gabor filters to capture rich facial movements in the mouth region that possibly occur in all directions.

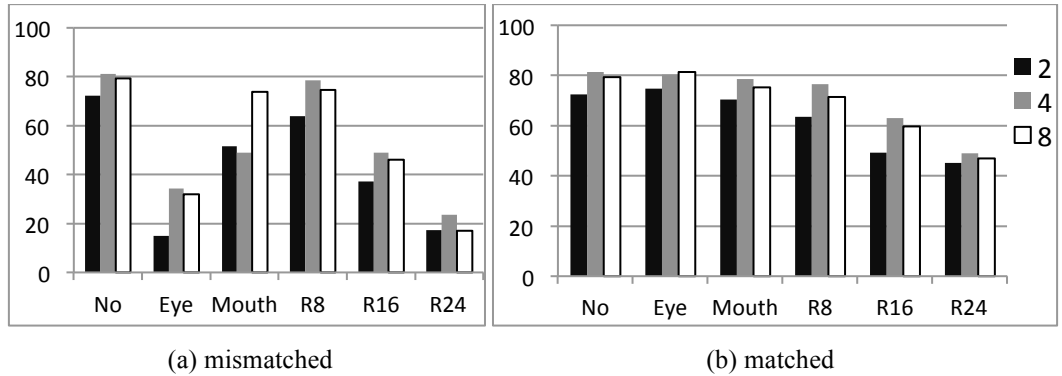


Fig. 9 Classification performance (%) obtained using three sets of orientation of Gabor filters (i.e. 2, 4 and 8) on the JAFFE database. The number of the scale of Gabor filters is fixed to eight.

Fig. 10 compares the accuracies obtained using four, six, eight, and ten scales of Gabor filters (the number of the orientation of Gabor filters is fixed to four). The four sets of scale are composed of $(5:2:11)$, $(5:2:15)$, $(5:2:19)$, and $(5:2:23)$ pixels respectively. Similar to the comparison results for different orientations of Gabor filters in Fig. 9, using eight scales of Gabor filters produces the highest accuracy among the four sets of scale values on nearly all types of occlusion under matched conditions, and it also performs the best for non-occlusion and occlusions of R_8 , R_{16} , and R_{32} under mismatched conditions. The results also confirm that the default four orientations and eight scales of Gabor filters are the optimal parameters for FER in occluded images, at least for non-occlusion and random occlusion. Two exceptions occur when the eyes and the mouth are occluded, for which the scales of six and four obtain the highest performance respectively under mismatched conditions. The results again imply the necessity to

set proper individual values to the orientation and scale of Gabor filters for FER in images with occluded eyes or mouth. It can be concluded that the changes in the orientation and scale of Gabor filters only pose limited impact on the performance of the proposed approach for most types of occlusion and both train-test strategies.

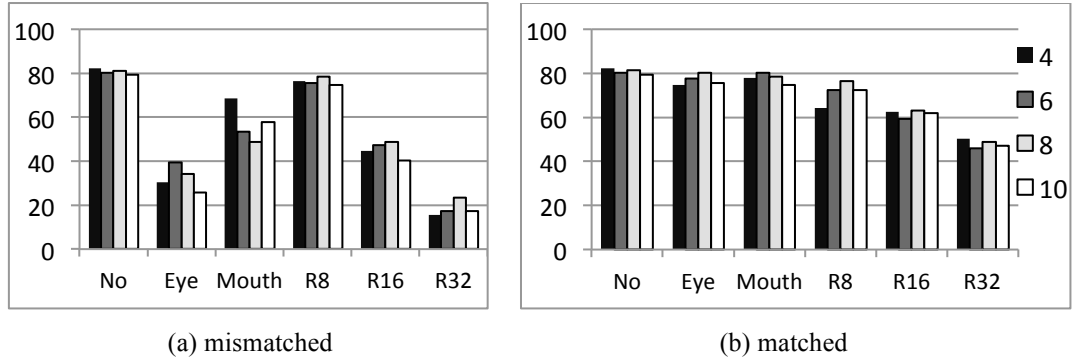
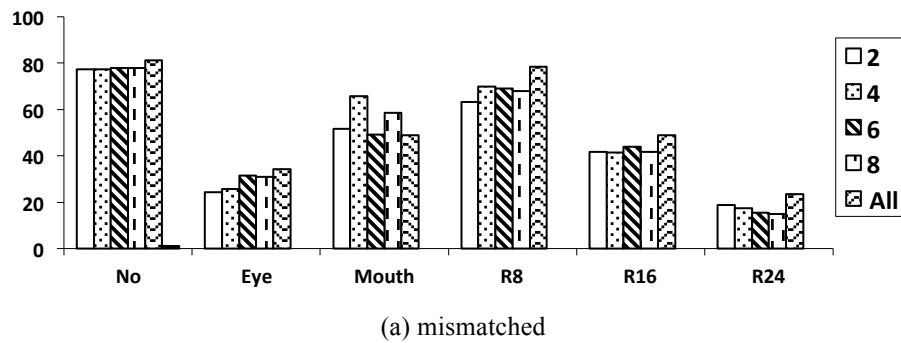


Fig. 10 Classification accuracy (%) obtained using four sets of scales of Gabor filters (i.e. 4, 6, 8 and 10) on the JAFFE database. The number of the orientation of Gabor filters is fixed to four.

(2) Template size

The size of the template determines the extent to which the useful discriminative features can be extracted from the facial image. Fig. 11 compares the accuracies obtained using four sets of template sizes (i.e. 2, 4, 6, and 8) individually and their combination (i.e. 'All'). It can be observed that all the four sets of template sizes have similar performance for non-occlusion and occlusion of R_8 , R_{16} , and R_{24} under both mismatched and matched conditions. Therefore, the changes in individual size of templates exert little influence on the performance. This is mainly due to the fact that these occlusions can be distributed equally at any position in the facial image, and thus the information for discriminating facial expressions also tends to be distributed equally in all sets of template sizes. The combination of all the four template sizes helps to achieve higher accuracy for all of the tested occlusion, except for occlusion of the mouth for which the sizes of four and eight perform the best under mismatched conditions and the size of eight the best under matched conditions. Thus, it seems to suggest that a combination of different sets of template sizes is advisable for FER on most types of occlusion in general, but for bigger occlusion at a roughly fixed location such as the mouth, a larger size of templates is recommended.



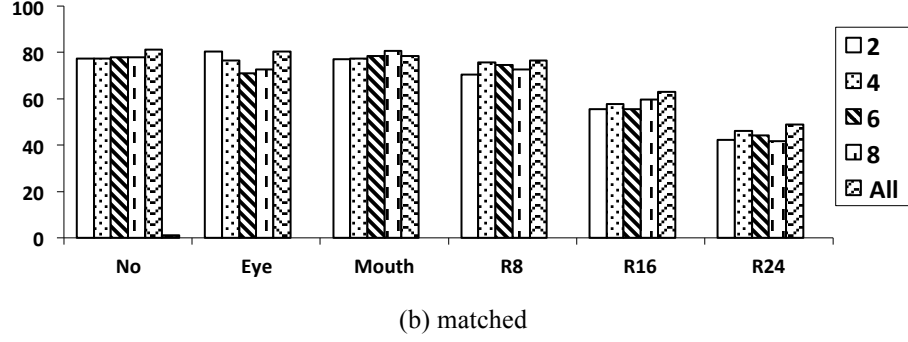


Fig. 11 Classification performance (%) obtained using four sets of template sizes (i.e. 2, 4, 6 and 8) and their combination on the JAFFE database.

(3) Occlusion ratio

Realistic occlusion may change dramatically in the ratio of its height and width. To simulate occlusions $R_{H \times W}$ with different ratios, we follow the process of simulating R_8 , R_{16} and R_{32} in Section 4.2, by adding a mask with a size of the height H and the width W into the face region at a random location. The height H is assumed to eight or 16 pixels for simplicity, but the width W is allowed to be a set of integer values within the range of (0 48), resulting in two sets of random occlusions with different ratios as shown in Fig. 12.

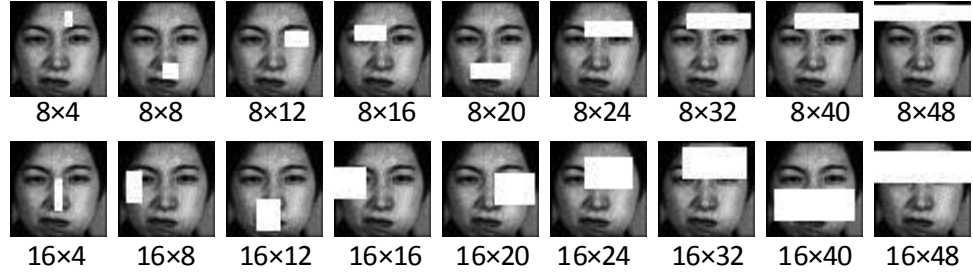


Fig. 12 Samples of two sets of simulated images with different occlusion ratios. ‘ $H \times W$ ’ indicates a simulated occlusion with a height H and a width W .

Fig. 13 shows the results of the proposed approach evaluated on the simulated occlusions with different ratios. As expected, the performance gradually decreases as the width of the occlusion increases under both mismatched and matched conditions, but the performance for occlusion with a height of 16 declines more quickly than that with a height of eight, and this is also the case for the performance under mismatched conditions as compared to under matched conditions. It is interesting to observe that, under matched conditions, the proposed approach has higher accuracy on $R_{16 \times 4}$ than on $R_{8 \times 4}$, and similar performance between $R_{16 \times 8}$ and $R_{8 \times 8}$. This is partly due to: a) performance fluctuation caused by the Monte Carlo algorithm for feature extraction; and b) occluded features on the left part of the face can be equivalently replaced by similar features on the right part, and vice versa (note that $R_{16 \times 4}$ and $R_{16 \times 8}$ appear more vertically than $R_{8 \times 4}$ and $R_{8 \times 8}$). However, the proposed approach still achieves around 60% and 40% accuracies on occlusion of $R_{8 \times 48}$ and $R_{16 \times 48}$ respectively under matched conditions, and around 40% and 20% accuracies respectively under mismatched conditions. In addition, the approach also exhibits a certain degree of robustness to changes in occlusion ratios, which is particularly

evidenced by the comparable performance obtained for all occlusion with the same height of eight and only 20% decline in accuracy when the width of the occlusion increases from 0 to 48 pixels under matched conditions.

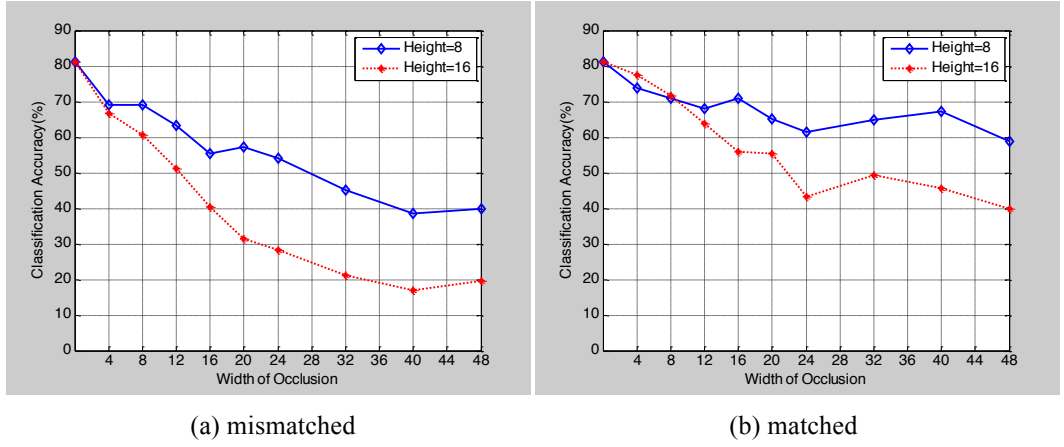


Fig. 13 Classification performance evaluated on two sets of random occlusion with different ratios on the JAFFE database. The two sets of simulated occlusion have the same height of 8 and 16 pixels respectively.

4.6 Classification performance on occlusion by glasses (eyewear)

To prove the robustness of the proposed approach, we also report the classification performance on two types of near-realistic occlusion by simulated clear and solid glasses as shown in Figure 5(b). Both matched and mismatched train-test strategies are used. For the mismatched strategy, in addition to using non-occluded images for training, we also give the results of using training images with patch-like occluded eyes. The adoption of occluded eyes is to test whether patch-like occlusion in training images helps to improve the classification performance for occlusion by glasses presented at the same location in testing images.

Table 2 shows the classification results for seven emotions using the proposed approach on the CK and JAFFE databases. The accuracy of the proposed approach reduces only by a small amount as a result of occlusion by clear or solid glasses, compared to results for no occlusion under matched train-test conditions. The reductions are 0.3% and 3.8% for occlusion by clear and solid glasses, respectively, on the CK database, and 0.4% and 6.1% on the JAFFE database. For a mismatched train-test strategy, the performance reduction remains small (<7.1%) on CK, but it becomes significant (>33%) on JAFFE. Compared with non-occluded training images, using training images with occluded eyes does not improve the performance on CK, but it leads to much better performance on JAFFE, particularly for occlusion by solid glasses. Occlusion by solid glasses leads to greater performance reduction than that by clear glasses for both the databases and this is true for both matched and mismatched train/test strategies. However, the performance reduction on CK is not as significant as that on JAFFE. As explained in Section 4.3, the reason for these contrasting results on the two databases lies in the fact that emotions in JAFFE are expressed mainly through the two eyes, while those in CK involve the mouth region. Thus, it appears that the impact of occlusion by glasses on FER performance is dependent on cultural differences in expression of emotion and hence on the database.

Table 2 Classification accuracy (%) of the proposed approach under occlusion by clear and solid glasses. ‘No’, ‘Clear’, ‘Solid’, and ‘Eye’ refer to no occlusion, occlusion by clear and solid glasses, and occlusion of the eyes, respectively.

	Matched (train-test)			Mismatched (train-test)			
	No-No	Clear-Clear	Solid-Solid	No-Clear	No-Solid	Eye-Clear	Eye-Solid
CK	95.3	95.0	91.5	94.4	89.7	89.5	88.2
JAFFE	81.2	79.8	75.1	42.7	20.7	47.9	35.2

4.7 Computational time performance

Table 3 displays the average computational time per image used for different processing steps at the training and the testing stages on the JAFFE database. The number (*Num*) of extracted templates for each emotion and each template size equals to 1000, and the distance metric is L_1 . The programs are developed in Matlab 7.6.0 with a dual core 3.20GHz CUP having 4GB RAM and a 64-bit Window 7 operating system. The program is quite fast and is able to process four testing images per second. Among all the processing steps, template matching is the most computationally expensive one, accounting for more than 70% of the overall time for both the training and testing stages. The training stage takes much longer than the testing stage (approximately 3.42 vs. 0.24 seconds), which is the preferred asymmetry because the training can be performed off-line. Due to the adoption of norm-based SVM for discriminative template selection, template matching at the testing stage takes as little as 6% of the time required by matching at the training stage.

Table 3 Average computational time (millisecond) per image on the JAFFE database (*Num* = 1000; L_1 distance)

	Gabor Feature	Template Extraction	Template Matching	Template Selection	SVM Classification	Overall Time
Training stage	54.3	15.8	2950	398.2	-	3418.3
Testing stage	54.3	15.8	173.2	-	1.0	244.3

4.8 Performance comparison with previous approaches

The performance of the proposed approach is compared with the reported results in previous approaches in this section. The approaches in Tables 3 and 4 are chosen due to their state-of-the-art performance on the CK and JAFFE databases, and they may adopt feature descriptors and emotion classifiers that are different from our approach. The performance is the average accuracy over 10 cross validations. For fair comparison, only the performance with occlusion of the eyes and occlusion of the mouth based on the *matched* strategy is used. The performance on non-occluded images is also given as a reference to evaluate the accuracy reduction caused by different types of occlusion. No previous research has reported tests on FER performance based on the *mismatched* strategy to the best of our knowledge and these results obtained by us are new and therefore not comparable to any previously reported.

Seen from the results on the CK database in Table 4, the proposed approach outperforms all benchmarked approaches in [9], [10] for both types of occlusion and no occlusion. The accuracy of the proposed approach is 0.8%, 2.8% and 3.6% higher for no occlusion, occluded eyes and occluded mouth, respectively, compared to the best performance in previous approaches. The reduction in accuracy owing to occlusion for our approach is the lowest (-0.2%) for occluded eyes and the second lowest (-4.5%) for occluded mouth.

Results on the JAFFE database in Table 5, show that the proposed approach has higher accuracy than the approaches in [7], [8], [16] and a comparable performance to the approaches in [10] for occlusion of the eyes and occlusion of the mouth. It should also be pointed out that the results in [7], [10], [16] are based on six basic emotions as opposed to seven emotions used in our approach. The higher performance in [10] over the proposed approach should be viewed taking this into consideration. The performance reductions (-0.9% and -2.8%) for our approach are the lowest among all compared approaches for occlusion of the eyes and occlusion of the mouth, respectively.

From the above analysis, we can see that the proposed approach not only achieves promising results for all three types of occlusion, but also shows robust performance to occluded eyes and mouth. The lower performance on JAFFE than that on CK is largely due to the fact that the expression of emotions in JAFFE images is more subtle than those in CK images. A shortcoming of the benchmarked approaches [9], [10] is that they extract a Gabor feature vector from facial regions without a further step of eliminating features influenced by occlusion. On the other hand, the proposed approach randomly extracts a set of templates from facial regions to represent local features and then utilizes SVM based feature selection to remove the templates influenced most by occlusion.

Table 4 Performance (%) Comparison on the CK Database

Reference	Feature	Classifier	No	Eyes	Mouth
Our	Gabor template	SVM	95.3	95.1 (-0.2)	90.8 (-4.5)
[9]	Gabor filter (Lfr)	MCC	94.5	92.3 (-2.2)	87.2 (-7.3)
	Gabor filter (Lfr)	CSM	93.6	91.5 (-2.1)	86.4 (-7.2)
[10]	Shape	SVM	91.4	88.4 (-3.0)	86.7 (-4.7)
	Gabor filter	NN	91.6	86.8 (-4.8)	84.4 (-7.2)
	DNMF	-	86.7	84.2 (-2.5)	82.9 (-3.8)

Note: ‘No’ stands for ‘no occlusion’, and negative values in parentheses indicate reductions in accuracy due to occlusion (with reference to no occlusion).

Table 5 Performance (%) Comparison on the JAFFE database

Reference	Feature	Classifier	No	Eyes	Mouth
Our	Gabor template	SVM	81.2	80.3 (-0.9)	78.4 (-2.8)
[7]	RPCA	Adaboost	87.5	68.8 (-18.7)	-
[8]	ALM+E	NN	-	77.7	73.3
	ALM+E	SVM	-	73.9	72.0
[16]	Motion units	Bayesian	70.3	67.1 (-3.2)	49.5 (-20.8)
[10]	Gabor filter	NN	88.1	83.1 (-5.0)	81.5 (-6.6)
	DNMF	-	85.2	82.5 (-2.7)	81.5 (-3.7)

5 Conclusion

This paper proposes a facial expression recognition approach robust to facial occlusion and investigates the effect of different types of occlusion on FER performance. We use a Monte Carlo algorithm to select Gabor templates from the gallery images and template matching over a search area to generate features that are robust to occlusion. This algorithm was tested with a range of different types of occlusion and found that the effect on the performance depends on the size of the occluded region. For 8x8 blocks the effect was seen to be minimal while for 24x24 blocks it is significant where the eye to eye distance is approximately 16. For train-test matched conditions, recognition accuracy is reduced less than 5% when the mouth or eyes are occluded. It is reduced only by 20% on the CK and 32% on the JAFFE database, respectively, even when a quarter of the face is occluded. Under near-realistic occlusion by simulated clear and solid glasses, the reductions are 0.3% and 3.8% accuracies respectively on the CK and 0.4% and 6.1% on the JAFFE. When the training images are not occluded but the test images are occluded (the mismatched condition), the methodology does not learn very well to reduce the effect of rectangular patch-like occlusion and hence the largest performance degradation is 65% on CK and 58% on JAFFE. However, the performance degradation is much lower for the more realistic eyewear occlusion cases, with less than 7.1% reduction for occlusion by both clear and solid glasses on the CK images, and around 33% reduction in accuracy for occlusion by solid glasses on the JAFFE images. The performance of the approach is little impacted by the changes in parameters such as the orientation and scale of Gabor filters and the size of templates, for most types of occlusion. The evaluations on occlusion of the eyes, the mouth and eyewear demonstrate that the proposed approach can achieve more than 75% and 90% accuracies on the JAFFE and CK databases respectively under matched conditions even when these traditionally recognized important regions for FER are occluded. Computational time tests on JAFFE images show that the proposed algorithm is fast and four test images can be processed per second on a dual core 3.2 GHz machine running Matlab.

The effect of occlusion can also be data dependent because of factors such as cultural differences in facial expression of emotions, differences in exaggeration of the facial expression by the subjects, etc., particularly evident in the mismatched train-test case. Occlusion of the mouth has a significant effect on all emotions except anger and happiness on CK images. Occlusion of the eyes has significant impact on anger, neutral, and surprise, while occlusion with 24x24 blocks has the biggest effect on fear, happiness and sadness on JAFFE images. For robust performance, it is best to avoid occlusion of the eyes or mouth and also use training data with the same type of occlusion as that expected to be present in test data. For the case that occluded mouth exists, our results on parameter sensitivity indicate that it is recommended to set larger values to the orientation and scale of Gabor filters and a larger size of templates for better classification performance. Our future work will evaluate the proposed approach on images with occlusion by real objects, such as sunglasses and scarves. The current exhaustive template matching can be made more time efficient by employing a “smarter” strategy to prune some of the matches, such as performing exhaustive template matching only in regions around the mouth and the two eyes.

References

1. Zeng Z, Pantic M, Roisman GI, Huang TS (2009) A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31 (1):39-58
2. Towner H, Slater M (2007) Reconstruction and Recognition of Occluded Facial Expressions Using PCA. In: *Affective Computing and Intelligent Interaction*. pp 36-47
3. Bourel F, Chibelushi, C.C., Low, A.A. Recognition of facial expressions in the presence of occlusion. In: *12th British Machine Vision Conference*, 2001. pp 213-222
4. Bourel F, Chibelushi CC, Low AA Robust facial expression recognition using a state-based model of spatially-localised facial dynamics. In: *Automatic Face and Gesture Recognition*, 2002. *Proceedings. Fifth IEEE International Conference on*, 2002. pp 106-111
5. Tan Dat N, Ranganath S Tracking facial features under occlusions and recognizing facial expressions in sign language. In: *Automatic Face & Gesture Recognition*, 2008. *FG '08. 8th IEEE International Conference on*, 2008. pp 1-7
6. Hammal Z, Arguin M, Gosselin F (2009) Comparing a novel model based on the transferable belief model with humans during the recognition of partially occluded facial expressions. *Journal of Vision* 9 (2):1-19
7. Xia M, YuLi X, Zheng L, Kang H, ShanWei L Robust facial expression recognition based on RPCA and AdaBoost. In: *Image Analysis for Multimedia Interactive Services*, 2009. *WIAMIS '09. 10th Workshop on*, 2009. pp 113-116
8. Jiang B, Jia K-b (2011) Research of Robust Facial Expression Recognition under Facial Occlusion Condition. In: *Active Media Technology*, vol 6890. *Lecture Notes in Computer Science*. Springer Berlin/Heidelberg, pp 92-100
9. Buciu I, Kotsia I, Pitas I Facial expression analysis under partial occlusion. In: *Acoustics, Speech, and Signal Processing*, 2005. *Proceedings. (ICASSP '05). IEEE International Conference on*, 2005. pp 453-456
10. Kotsia I, Buciu I, Pitas I (2008) An analysis of facial expression recognition under partial facial image occlusion. *Image and Vision Computing* 26 (7):1052-1067
11. Bassili JN (1979) Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology* 37 (11):2049-2058
12. Nusseck M, Cunningham DW, Wallraven C, Bülthoff HH (2008) The contribution of different facial regions to the recognition of conversational expressions. *Journal of Vision* 8(8):1, 1-23
13. Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007) Robust Object Recognition with Cortex-Like Mechanisms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29 (3):411-426
14. Jhuang H, Serre T, Wolf L, Poggio T A Biologically Inspired System for Action Recognition. In: *Computer Vision*, 2007. *ICCV 2007. IEEE 11th International Conference on*, 2007. pp 1-8
15. Ligang Z, Tjondronegoro D, Chandran V Toward a more robust facial expression recognition in occluded images using randomly sampled Gabor based templates. In: *Multimedia and Expo (ICME)*, 2011 *IEEE International Conference on*, 11-15 July 2011. pp 1-6
16. Miyakoshi Y, Kato S Facial emotion detection considering partial occlusion of face using Bayesian network. In: *Computers & Informatics (ISCI)*, 2011 *IEEE Symposium on*, 20-23 March 2011. pp 96-101.
17. Cotter SF Sparse Representation for accurate classification of corrupted and occluded facial expressions. In: *Acoustics Speech and Signal Processing (ICASSP)*, 2010 *IEEE International Conference on*, 14-19 March 2010. pp 838-841
18. Cotter SF Recognition of occluded facial expressions using a Fusion of Localized Sparse Representation Classifiers. In: *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, 2011 *IEEE*, 4-7 Jan. 2011. pp 437-442

19. Yongmian Z, Qiang J (2005) Active and dynamic information fusion for facial expression understanding from image sequences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27 (5):699-714
20. Huang X, Zhao G, Zheng W, Pietikäinen M (2012) Towards a dynamic expression recognition system under facial occlusion. *Pattern Recognition Letters* 33 (16):2181-2191
21. Halliday LA (2008) Emotion detection: can perceivers identify an emotion from limited information? Paper presented at the M.S. thesis, University of Canterbury
22. Gabor D (1946) Theory of communication. *Institution of Electrical Engineers -- Journal -- Radio and Communication Engineering* 93 (26, Part III):429-457
23. Dunja M, Janez B, Marko G, Natasa M-F Feature selection using linear classifier weights: interaction with classification models. In: 27th annual international ACM SIGIR conference on research and development in information retrieval, Sheffield, United Kingdom, 2004. ACM, pp 234-241
24. Kanade T, Cohn JF, Yingli T Comprehensive database for facial expression analysis. In: *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000. pp 46-53
25. Lyons M, Akamatsu S, Kamachi M, Gyoba J Coding facial expressions with Gabor wavelets. In: *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998. pp 200-205
26. Jongsun K, Jongmoo C, Juneho Y, Turk M (2005) Effective representation using ICA for face recognition robust to local distortion and partial occlusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27 (12):1977-1981
27. Ligang Z, Tjondronegoro D (2011) Facial Expression Recognition Using Facial Movement Features. *Affective Computing, IEEE Transactions on* 2 (4):219-229